Structure and Organization of the Human Transglutaminase 3 Gene: Evolutionary Relationship to the Transglutaminase Family

In-Gyu Kim, Seung-Chul Lee, Jeung-Hoon Lee, Jun-Mo Yang, Soo-Il Chung,* and Peter M. Steinert

Skin Biology Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases; and *Laboratory of Cellular Oncology and Development, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland, U.S.A.

The human haploid genome contains a family of at least five different transglutaminases that are differentially expressed in time- and tissue-specific ways. Of these, transglutaminase 3 (TGase3) is unusual in that it is a pro-enzyme requiring activation by proteolysis. To date it is known to be expressed only in terminally differentiating epidermal and hair follicle keratinocytes. In this paper we show that it is encoded by a gene (TGM3) of 42.8 kbp containing 13 exons. In the course of isolation of genomic clones for the TGM3 gene, we also found clones encoding the widely expressed tissue or TGase2 enzyme, perhaps due to high degrees of sequence homology. The structure of the TGM2 gene has not yet been reported. Our incomplete data suggest its exon/intron organization is very similar to that of TGM3. Although the common intron splice points of all members of the transglutaminase gene family have been conserved, the TGM3 and TGM2 genes, and the gene for the subplasma membrane transglutaminase-like protein band 4.2, lack two introns found in the TGM1 and factor XIIIa genes, and the exact intron splice point of another intron is shifted with respect to that of the TGM1 and factor XIIIa genes. Based on sequence homologies and gene structures, the data support a phylogenic tree in which the TGM2 and TGM3 genes belong on a branch distinct from other transglutaminases. *Key words: transglutaminase 2/ cell envelope/keratinization. J Invest Dermatol 103:137–142, 1994*

Transglutaminases (TGases; protein-glutamine:amine γ-glutamyltransferases, EC 2.3.2.13) catalyze the formation of an $N^{\epsilon}(\gamma$-glutamyl)lysine isodipeptide cross-link in proteins between the γ-amide of a donor glutamine and the ϵ-NH_2 of an acceptor lysine. The result is a stable, insoluble macromolecular structure, a process used widely throughout the plant and animal kingdoms [1–5]. In humans, a family of five distinct TGases have been described to date (reviewed recently in [4–6]), including the catalytic a subunit of factor XIII; band 4.2, an inactive enzyme that forms a part of the subplasma membrane of animal cells; TGase1, a membrane-associated enzyme present in many epithelial as well as some non-epithelial tissues; a ubiquitously expressed tissue TGase2; and a proezyme activity, TGase3, so far found only in terminally differentiating epidermal and hair keratinocytes. Recently, a TGase4 activity was described in rat prostate [7], but it is not yet known whether it is also present in humans. All of these proteins show remarkable conservation of amino acid sequences and likely secondary structures, especially in the central two thirds of the molecule [6], a region that apparently describes their common catalytic properties. The enzymes differ from each other primarily in their amino- and carboxyl-termini, for example, in factor XIIIa [8,9] and TGase1 [6,10–12], by recruitment of additional sequences that connote specific properties and functions. The TGase3 is different in the sense that it is a pro-enzyme, requiring activation by proteolytic cleavage at a specific site about two thirds of the way long the molecule defined by a flexible hinge that is unique in the TGase family [13,14].

Not surprisingly, the known structures of the genes for three of the five human TGases have also been conserved in structure, thus indicating that the genes for the known extant family members arose from a common ancestor. The locations of most intron splice points have been conserved, although in comparison to the gene for band 4.2, the TGase1 and factor XIIIa genes possess two additional introns, one in 5′-untranslated regions and a second about 60% of the way along the coding sequences.

However, less is known about the precise function(s) of these enzymes. Interestingly, the three active enzymes TGase1, TGase2, and TGase3 are expressed in the epidermis [14]. Whereas the ubiquitous tissue or TGase2 activity is thought to be involved in apoptosis [15–17], the TGase1 and TGase3 enzymes are thought to be involved in the construction of a cornified cell envelope during terminal differentiation that provides a vital barrier function for the tissue [18–20]. Expression data show that although the proTGase3 mRNA is only about 2% as abundant as that for TGase1 [14], activated TGase3 nevertheless accounts for up to 75% of the total TGase activity in mammalian epidermis [13]. Accordingly, the characterization of the cDNA, protein, and gene structures of these three TGases is essential to understand their role(s) in the processes of normal and abnormal differentiation of the epidermis. We have recently described the full-length cDNA structure of the TGase3 system [14] and now describe in this paper its gene structure. During the isolation of genomic clones encoding the TGM3 gene, we also encountered clones encoding much of the TGM2 gene, and report here that both its gene structures are very similar to that of the

Manuscript received February 8, 1994; accepted for publication April 5, 1994.

Dr. Kim's present address: Department of Biochemistry, Inha University Medical School, Inchon, Republic of Korea.

Reprint requests to: Dr. Peter M. Steinert, National Institutes of Health, Building 6, Room 425, Bethesda, MD 20892.

Abbreviations: bp, base pair; kbp, kilobase pairs; nt, nucleotide(s); TGase, transglutaminase.

0022-202X/94/$07.00

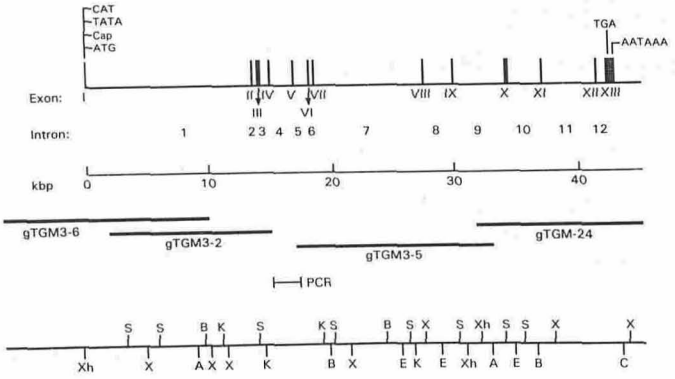

Figure 1. Organization of the human TGM3 gene. The *upper* line represents a schematic diagram of the gene showing the organization of exons (I–XIII) and introns (1–12). The positions of the likely regulatory sequences are illustrated. The *second* line denotes size scale. The *third* line illustrates the four overlapping genomic clones used to assemble the gene structure. The 2-kbp PCR fragment that joins the clones gTGM3-2 and gTGM3-5 is shown. The *lower* line illustrates representative restriction enzyme sites used for the construction of the above maps. A, *Aat* II; B, *Bam* HI; C, *Cla* I; E, *Eco* RI; K, *Kpn* I; S, *Sph* I; X, *Xba* I; and Xh, *Xho* I. Diagnostic *Xho* I and *Aat* II sites confirmed the overlap of clones gTGM3-6/gTGM3-2 and gTGM3-5/gTGM3-24.

TGM3 system. The evolutionary significance of this finding is discussed.

MATERIALS AND METHODS

Molecular Biology Procedures A human placental genomic library constructed in the λ-phage EMBL-3 vector (Clontech) was screened with a series of 15 synthetic oligonucleotides of 50 nt in length, corresponding to predicted exons, based on our full-length cDNA sequence for TGase3 [6,14]. A total of 12 genomic clones were thus identified and plaque purified. Each full-length insert was excised, subcloned into the pGEM-7z vector (Promega) and subjected to further restriction enzyme mapping. Primers for DNA sequencing with the Sequenase 2.0 system (U.S. Biochemical Corp.) on both strands were located in the predicted exon sequences to cross over exon-intron boundaries. Following these mapping and sequencing studies, we found that the available genomic clones did not encode exon I. Accordingly, we screened a different EMBL-3 human genomic library (lymphocyte, Clontech), and isolated another genomic clone as above that contained exon I and sequences upstream of the gene. A likely cap-site of the TGM3 gene was established in an RNA-mediated anchored polymerase chain reaction (PCR) and characterization of the resulting sequence information [21]. In addition, our available genomic clones did not contain exon V sequences. The gap encompassing exon V was closed by use of the PCR using as primers exon IV sequences at the 3′-end of clone gTGM3-2 and intron V sequenes at the 5′-end of clone gTGM3-5 (see below and Fig 1). All PCR reactions were done with the Perkin Elmer-Cetus amplification kit using 25 pmol of primers and with conditions of 95°C (5 min), and 35 cycles of denaturation at 94°C (0.5 min), annealing at 42°C (0.5 min), and elongation at 72°C (1.5 min). The products were fractionated through a 4% low-melting agarose gel, excised, and purified through Chroma spin 100 columns (Clontech). The ends of the amplified DNA were filled in with the

Table I. Sizes of Exons and Introns (in bp) of the Human TGM3 Gene

Exon		Intron		
Number	Size	Number	Size[a]	Method
I	48[b]	1	13,400	PCR/mapping
II	174	2	313	Sequencing
III	240	3	600	PCR
IV	119	4	1,800	PCR
V	129	5	1,300	PCR
VI	178	6	103	Sequencing
VII	136	7	8,600	Mapping
VIII	104	8	2,200	PCR
IX	146	9	4,000	PCR/mapping
X	309	10	3,300	PCR/mapping
XI	158	11	4,000	PCR/mapping
XII	134	12	500	PCR
XIII	645[c]			

[a] Sizes of introns are estimated to be within 100 bp (except those determined exactly by sequencing).
[b] Includes 44 bp of sequences from likely cap-site shown in Fig 2 to initiation codon plus 4 bp of coding sequences.
[c] Includes from termination codon to polyadenylation signal sequence.

Table II. Splice Donor and Acceptor Sequences Used in the Human TGM3 Gene[a]

Donor Sequence	Intron Number	Acceptor Sequence
* A A		L G V Q
CGAAACATGGCTGgtgagtgcatggcatcttctccatcaggtc	1	taggacttcaggttcgccttctgcatgcagCTCTAGGAGTCCAG
I V S T G		P Y P S
ATTGTGTCCACAGgtacctgctcattccccctccttgcccaaac	2	gttggttttctcaacctctgtctttgacagGGCCTTACCCCTCA
P W L N V		D S V F
CCCTGGCTGAATGgtaggtgtctagccacccacactctcagcc	3	tcgctggtgtctgctttttgtatcaaatagTGGATAGCGTCTTT
N F G Q		F E E D I
GAACTTTGGACAGgtaaaatcataaggaataccatctccacca	4	aattgtaatgtatggtctcgttctgaacagTTTGAAGAAGACAT
L S A M		I N S N D
GCTGAGTGCCATGgtgagtaacatggtcaatgctgtcaggctg	5	ctcattttgggggggtggtttctggcccagATCAATAGCAATGA
T L N T A		L R S L
ACCCTCAACACAGgtaccttgggtgtggtgtgccttggctggg	6	ctgagtcctcacgcccaccctgacttgcagCGCTGCGGTCTTTG
D S V W		N F H V W
GTGATAGCGTATGgtaagtatctcacctttttccctgaacttcg	7	aaatgagtgggacacagtcaggggctgaagGAATTTCCATGTCT
E R S Q G		V F Q C
GAAAGAAGCCAAGgtaacttctctgggtgtggttctgagtcgt	8	attccagcctctgcgcatcttggcctccagGGGTGTTCCAGTGC
K Y P E G		S D Q E
AAGTACCCAGAAGgtaggagggacgctggcggggcagtgcgcg	9	gcatgtttgtcgtcttttcccaccacacagGCTCTGACCAGGAA
D P E E E		A E H P
GACCCTGAGGAAGgtaacgcatcccgcagttggaggagatcca	10	tcactcggatcccctggcttctccttccagAGGCAGAACATCCC
L T L E		V L N E A
CTTGACCCTGGAGgtaatggggttcccccatcctgtgggaagg	11	tcctgactcggcagcccctctcccccatagGTGCTGAACGAGGC
L K I D		V P T L G
ACCTGAAGATCGAgtgagtcctgggcctaagtggcggtgcagg	12	atcaacactgatctgtgcccctcccccatcagCGTGCCGACCCTAG

[a] The sequences of the 5′-splice donor (left column) and 3′-splice acceptor sites are shown for each intron. Coding sequences are in capital letters; intron sequences are in lower case letters. *, initiation codon.

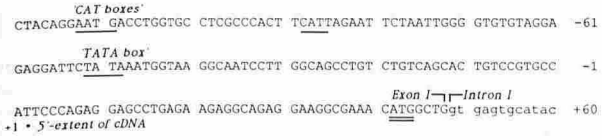

Figure 2. Structure of the 5′-end of the human TGM3 gene. The sequence of the genomic clone g-TGM3-6 immediately upstream of the likely initiation codon was determined and found to be almost identical to the available cDNA information adduced previously [14]. By both computer predictions and RNA-mediated anchored-PCR primer extension experiments, the indicated cap-site is the most likely site of initiation of transcription. Nearby likely TATA and CAT boxes identified in genomic sequences are *underlined*. The initiation codon is *double underlined*. The 5′-extent of the cDNA information [14] is shown, as is also the splice point of intron 1.

Klenow DNA polymerase, and subcloned into the pGEM-3z vector for sequencing [14].

Computer Analyses of Sequences Nucleic acid and protein sequence homologies were performed using the University of Wisconsin software packages compiled by the Wisconsin Genetics Computer Group [22] and the IBI Pustell sequence (version 3.5, International Biotechnologies Inc) [23,24] or Geneworks (Intelligenics) [25] software. A phylogenic tree was established as described [26].

RESULTS AND DISCUSSION

Isolation of Genomic Clones Encoding the TGM3 Gene Recent data have established that the locations of the exon/intron boundaries of the genes for three of five members of the human TGase family have been largely conserved [6,10,11,27,28]. Based on the assumption that the structure of the TGM3 gene is also similar, a series of synthetic oligonucleotides 50 nt long corresponding to predicted exon regions was used to probe a human placental genomic library, from which 12 genomic clones were isolated and plaque purified. Each cross-reacted with one or more of the synthetic oligonucleotides, except the probes designed for exons I and V. Each genomic clone was subjected to further restriction enzyme mapping and sequencing across the predicted exon/intron boundaries. Several clones were found to be identical. In this way, it was established that most of the TGM3 gene was encoded by the three genomic clones in 5′ to 3′ order of gTGM3-2 (14.2 kbp), gTGM3-5 (16.0 kbp), and gTGM3-24 (13.2 kbp) (Fig 1). Confirmation of a 1.8-kbp overlap between gTGM3-5 and gTGM3-24 was established by both restriction enzyme mapping and by hybridizations of other additional oligonucleotides located at the 3′ and 5′ ends of these clones, respectively. However, it was clear that a gap existed between clones gTGM3-2 and gTGM3-5, encompassing likely exon V sequences. This gap, in intron sequences, was found to be only 1.8 kbp long by use of a PCR experiment, using primers located at the 3′ and 5′ ends of these two clones, and total genomic DNA isolated from the human placental library as template. Subsequent hybridization and direct sequencing established that exon V sequences were present in the amplified fragment. The 5′ extent of the clone gTGM3-2 began in intron 1 sequences and did not contain exon I. Accordingly, a different human (lymphocyte) genomic library was screened. The 18-kbp clone gTGM3-6 was isolated, characterized and found by PCR and restriction enzyme mapping to overlap gTGM3-2 by about 10 kbp (Fig 1). By PCR analysis, exon I sequences were located on gTGM3-6 about 2.2 kbp above the 5′ extent of gTGM3-2, so that intron 1 is about 13.4 kbp long (Table I). Thus gTGM3-6 contains about 6 kbp of flanking sequences above the likely cap site of the TGM3 gene (Fig 1).

Accordingly, it was possible to assemble the full-length gene by this combination of direct sequencing, sequencing across exon/intron boundaries, and restriction enzyme mapping. In most cases, a close approximation of the sizes of introns was confirmed by additional PCR experiments. The sizes of the longest introns 1 and 6 could be estimated only from the sizes of restriction enzyme fragments and/or PCR products. The TGM3 gene consists of 13 exons and 12 introns, as summarized in Table I. All of the junctions conform to consensus splice donor/acceptor boundaries [29] (Table II).

During the course of this work, we resequenced the entire coding portion of the human TGM3 gene, and found four sequence differences with respect to the cDNA sequence [14], in codons 12 (AAA to ACA, lysine to threonine); 260 (AAA to ACA, lysine to threonine); 264 (TTC to CTC, phenylalanine to leucine); and 299 (GCT to CCT, alanine to proline). These are most likely simple sequence polymorphisms, for the following reasons. First, even though the entire cDNA sequences had been adduced by repeated PCR procedures and required resolution of seven ambiguous nt [14], none of the variations found in this study had been seen in the previous replicate PCR experiments. Second, we showed that the cDNA sequences directed the expression of an active TGase3 pro-enzyme protein in yeast [14]. Each of these amino acid substitutions occurs in regions predicted to form protein turns or Ω loops [6,22]. Thus the variations observed here in the TGM3 gene are polymorphisms in the human population that do not affect enzyme function or activity.

When compared with the available cDNA information, intron 1

Table III. Known Splice Donor and Acceptor Sequences Used in the Human TGM2 Gene[a]

Donor Sequence	Intron Number	Acceptor Sequence
* A E E		L V L E
ACCATGGCCGAGGgtcgccacactcccaggtaccatagttcta	1	ggctcatgcgtctccttctgttttccatagAGCTGGTCTTAGA
V V T G		P A P S
GTGTCGTGACCGGgtgagtacctgcaaccccactccaccacca	2	tgctcacggctactgcttctccctccacagGCCAGCCCCTAGC
A W C P A		D A V Y
GCCTGGTGCCCAGgtcctccactggctaccagggatccagctt	3	cagagctgaggtcccttttcttcctgcctagCGGATGCTGTGTA
N F G Q		F Q D G I
GAATTTTGGGCAGgtagggccacaccctgcctctcccaaccat	4	nnnnnnnnnnTTTCAAGATGGGA
G S A M		V N C N C
GGGTAGTGCCATGnnnnnnnnnn	5	ctcgtccccgtgctgccccactggttgcagGTCAACTGCAACT
V A C T V		* L R C L
GTGGCCTGCACAGgtgagctgcagctgggatgtgggtcatgag	6	ctcaccccgctttccactcctctccaacagTGCTGAGGTGCCT
E M I W		N F H C
GCGAGATGATCTGgtgaggtgggccagggtgggacggagccca	7	tgcttcattcccgccccccatcccccctcagGAACTTCCACTGC
E K S E G		T Y C C
GAGAAGAGCGAAGgtgcgtgggcgggcactgtcgcgtcaaaca	8	tcccgcccttcccctttctacctctgccagGAACGTACTGCTG
K Y P E		
CAAATACCCAGAGgatgttgccctcagggtcttaagcagcctt	9	Not known

[a] The sequences of the 5′-splice donor *(left column)* and 3′-splice acceptor sites are shown for each intron. Coding sequences are in capital letters; intron sequences are in lower case letters. *, initiation codon. n, unknown intron sequences around expected exon V. Intron splice points for expected exons V and X to XIII are not known, due to unavailable gene clone information.

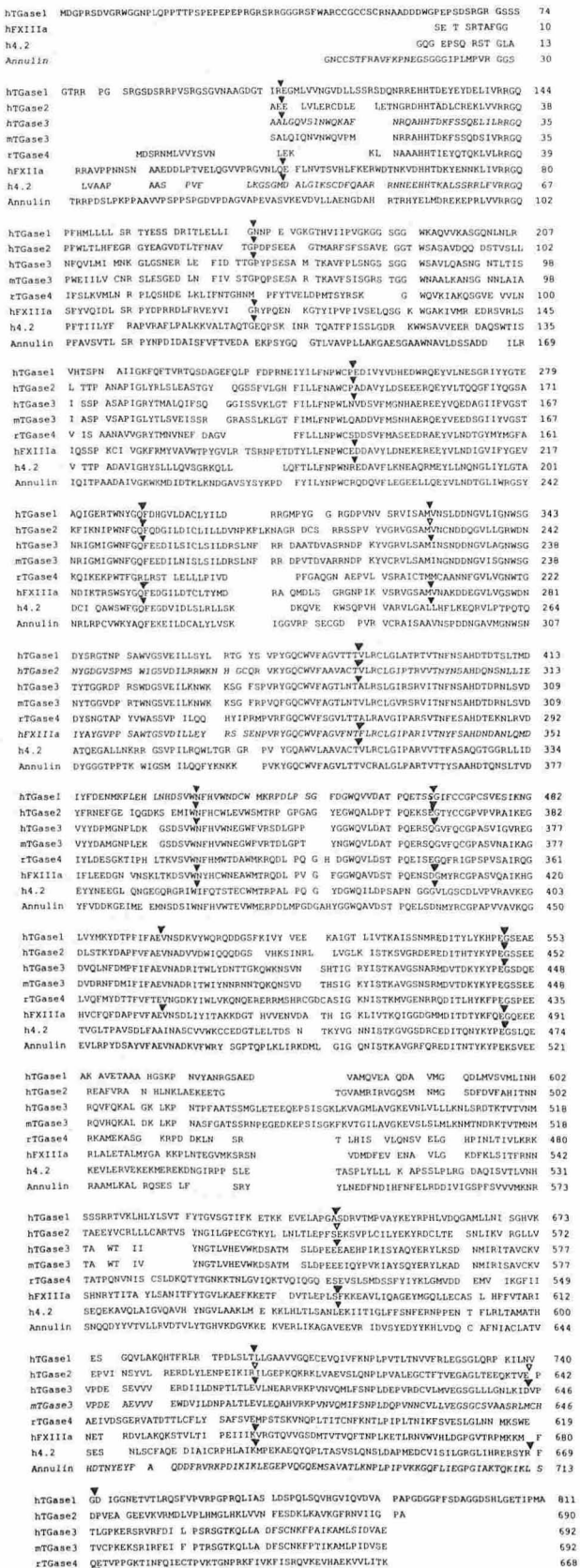

Figure 3. Comparison of the gene structures of several TGase-like proteins. Protein sequences are aligned to maximize homologies [25]. The known positions of the intron splice points in the genes are designated by *closed arrowheads*. The predicted locations of the other intron splice points in the TGM2 gene are designated by *open arrowheads*. h, human; m, mouse; hF, human factor XIIIa; h4.2, human band 4.2.

splices 4 bp downstream from the consensus initiation codon, so that exon I consists mostly of 5′-untranslated sequences (Fig 2). The genomic and available cDNA sequences differed by only one nt in the 5′-untranslated region. Further sequencing on the genomic clone upstream revealed a likely cap-site 1 nt above the available cDNA information, as predicted by computer programs. The available cDNA information had been adduced by use of RNA-mediated anchored-PCR primer-extension experiments [14]. In an attempt to confirm this potential cap-site, we performed additional experiments with different batches of foreskin epidermal RNA, and using primers located at the confirmed beginning of the exon II sequences. We obtained amplified products that extended to the same place, within 1 nt of the indicated cap-site of Fig 2, so that exon I is likely to be only 48 bp in length. Attempts to rigorously prove this as a functional cap-site by use of RNAse protection experiments with human foreskin RNA with a ^{32}P-labeled 4.3-kbp *Xho* I–*Sph* I fragment were unsuccessful. We think the most likely reason for this is the very low abundance of the mRNA encoding the TGase3 in the epidermis [14], and because the resulting exon I is only 48 bp, perhaps too short to confidently prove by this method [21,30]. However, in further good support of our model, a consensus TATA box exists at −52 nt and two consensus CAT boxies exist at −89 nt and −113 nt (opposite strand) (Fig 2). Additional functional assays involving reporter gene elements will be necessary to prove this model.

Altogether, these present data suggest that the human TGM3 gene is approximately 42.8 kbp long (Table I), about twice the length of the gene for the band 4.2 protein [27], three times longer than the TGM1 gene [6], but much shorter than the gene for the factor XIIIa [28].

Discovery and Partial Characterization of Genomic Clones for the TGM2 Gene However, three other genomic clones, which hybridized strongly to oligonucleotides designed from our known full-length cDNA for TGase3, clearly did not encode the TGM3 gene. Sequence comparisons established that they encoded TGase2 sequences instead [27]. The TGM2 gene structure has not been reported yet. Partial restriction enzyme mapping analyses indicate that most of the TGM2 gene is encoded by the three clones gTGM2-7 (13 kbp, containing exons I, II, III, and IV), gTGM2-10 (15 kbp, containing exons VI, VII, and VIII), and gTGM2-3 (13 kbp, containing exon IX). We were unable to close a gap between clones gTGM2-7 and gTGM2-10 containing exon V sequences by use of PCR methods, as described above for the TGM3 gene, perhaps suggesting that introns 4 and/or 5 are much larger than in the TGM3 gene. Nor does the clone gTGM2-3 contain expected exons X–XIII. Comparisons of the sequence of the clone gTGM2-7 with the available cDNA information encoding TGase2 [27] indicates that the TGM2 gene does not contain an intron in 5′-untranslated sequences. Sequencing across all of the available exon/intron boundaries reveals (Table III) that the introns splice at positions that are conserved with respect to the TGM3 gene. Thus although the exons are of similar size, it seems likely that the TGM2 gene is larger than the TGM1 or TGM3 genes due to relatively larger introns. Also, all of the known junctions again conform to consensus splice donor/acceptor boundaries (Table III).

Conservation of the Structures of the Genes of the Transglutaminase Family Figure 3 shows an updated homology alignment of available TGase proteins, and includes the newly described rat prostate TGase4 enzyme [7] and an invertebrate TGase protein from the grasshopper, termed annulin [31]. Several points are note-

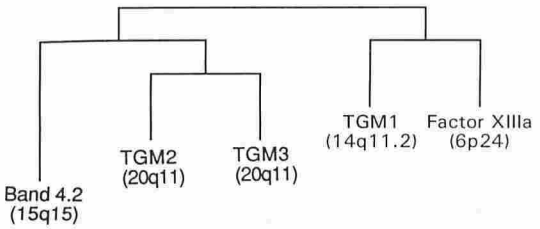

Figure 4. Phylogenic tree of the known human TGases. Using the neighbor-joining method [26] one possible rooted tree is illustrated, and is based on amino acid sequence homology scores and gene structures. Our new data confirm and extend an earlier hypothesis [11]. Two main branches are postulated, one including the genes for TGM1 and factor XIIIa, and a second containing the genes for band 4.2, TGM2, and TGM3. Because of their high degrees of homology, it is thought that the genes for TGM2 and TGM3 duplicated more recently from an ancestor in common with band 4.2. This view seems to be supported by the chromosomal locations of the five genes (shown in parentheses).

worthy. First, most intron positions have been precisely conserved. Second, the TGM1 and factor XIIIa genes have two additional introns: one in 5′-untranslated sequences; and another about 60% of the way along the coding sequences that splices exon VIII of the TGM2, TGM3, and band 4.2 genes. In addition, although the intron 12 splice point of the TGM3 gene is located in the same place as that of the gene for band 4.2, it is 6 nt upstream of the position of the corresponding intron splice point in the TGM1 and factor XIIIa genes.

Furthermore, as pointed out earlier [6], sequence and likely secondary structures between the regions spanned by introns 1 and 11 (for TGM2, TGM3, and band 4.2 genes) or 2 and 13 (for TGM1 and factor XIIIa genes) have been conserved, not only in all of the mammalian TGases but also in the invertebrate protein annulin. Homology scores using existing algorithms [24,25] suggest the TGase3 protein is closely related to TGase2 and TGase1, and more distantly related to factor XIIIa and band 4.2. Similarly, the rat TGase4 protein shows higher sequence homology to the human TGase1 and factor XIIIa proteins. TGase3 differs from all other members of the TGase gene family primarily by the net insertion of 12 residues that serve as a flexible hinge at its site of proteolytic activation (Fig 3) [14].

Evolution of the TGM3 Gene System The above new data on the likely structures of the TGM3 and TGM2 genes, together with their high degrees of sequence homology, allow a model for the evolution of the family of human TGases [26]. These data support and extend the notion [11] that an early or primordial gene diverged into two branches (Fig 4). One branch contains the TGM1 and factor XIIIa genes that have acquired an additional exon of sequences on their 5′-ends to connote specialized properties of membrane association (for TGase1 [32]) or proteolytic activation (for factor XIIIa [8,9]), and have either gained or retained an intron at about 60% along the coding sequences. A second branch includes the genes for band 4.2, TGM2 and TGM3 that may have never recruited an extra exon in 5′-untranslated sequences, may have lost an intron 60% along the coding sequences, and at least two members, possibly also the TGM2 gene, have moved the last intron splice point. We suggest that the TGM2 and TGM3 genes may be more modern members of the human TGase family in the sense that they have more recently duplicated, so that their organization and coding sequences are still highly homologous. This map is also consistent with the chromosomal locations of the TGase genes (Fig 4). Preliminary data† have shown that the TGM2 and TGM3 genes are tightly linked and map to chromosome 20. In contrast, the genes for factor XIIIa, band 4.2, and TGM1 have been scattered on the human genome, to chromosome positions 6p24–34 [33], 15q15 [34], and 14q11.2 [6,12], respectively, thereby supporting the view that the TGM2 and TGM3 genes have most recently duplicated and have not yet scattered on the genome. The gene for band 4.2 has a lower overall sequence homology, but this may be because it is not a functional enzyme and so there has been less selective pressure for sequence conservation [11]. Placement of TGase4 on the phylogenic tree of Fig 4 should await determination of its gene structure; that is, whether or not it contains the two additional introns and one displaced intron splice point common to the TGM1 and factor XIIIa genes.

In conclusion, the present data describe for the first time the organization of the TGM3 gene, and establish it to be most similar in terms of structure to the TGM2 gene, not the three other presently known human TGase genes.

† Wang M, Kim I-G, Steinert P, McBride OW (unpublished observations).

We are especially indebted to Mr. George Poy for the synthesis of all of the oligonucleotides used in this work. We are also grateful to Drs. Jack Folk and Sang-Chul Park for their friendship, advice and support throughout this work.

REFERENCES

1. Folk JE, Finlayson JS: The ϵ-(γ-glutamyl)lysine crosslink and the catalytic role of transglutaminases. *Adv Protein Chem* 31:1–133, 1977
2. Folk JE: Mechanism and basis for specificity of transglutaminase-catalyzed ϵ-(γ-glutamyl)lysine bond formation. *Adv Enzymol* 54:1–56, 1983
3. Lorand L: Transglutaminase-mediated crosslinking of proteins and cell aging: the erythrocyte and lens models. In: Zappia V, Galletti P, Porta R, Wold F (eds.). *Advances in Post-Translational Modifications of Proteins and Aging, Vol 231.* Plenum Press, New York, 1988, pp 79–94
4. Greenberg CS, Birckbichler PJ, Rice RH: Transglutaminases: multifunctional crosslinking enzymes that stabilize tissues. *FASEB J* 5:3071–3077, 1991
5. Polakowska RR, Goldsmith LA: The cell envelope and transglutaminases. In: Goldsmith LA (ed.). *Physiology, Biochemistry and Molecular Biology of the Skin.* Oxford University Press, New York, 1991, pp 168–201
6. Kim I-G, McBride OW, Wang M, Kim S-Y, Idler WW, Steinert PM: Structure and organization of the human transglutaminase 1 (TGM1) gene. *J Biol Chem* 267:7710–7717, 1992
7. Ho K-C, Quarmby VE, French FS, Wilson EM: Molecular cloning of the rat prostate transglutaminase complementary DNA. The major androgen-regulated protein DP 1 of rat dorsal prostate and coagulating gland. *J Biol Chem* 267:12660–12667, 1992
8. Ichinose A, McMullen BA, Fujikawa K, Davie EW: Amino acid sequence of the a subunit of human factor XIII. *Biochemistry* 25:4633–4638, 1986
9. Takahashi N, Takahashi Y, Putman FW: Primary structure of blood coagulation factor XIIIa (fibrinoligase, transglutaminase) from human placenta. *Proc Natl Acad Sci USA* 83:8019–8023, 1986
10. Phillips MA, Stewart BE, Rice RH: Genomic structure of keratinocyte transglutaminase. Recruitment of a new exon for modified function. *J Biol Chem* 267:2282–2286, 1992
11. Polakowska RR, Eickbush T, Falciano V, Razvi R, Goldsmith LA: Organization and evolution of the human epidermal keratinocyte transglutaminase 1 gene. *Proc Natl Acad Sci USA* 89:4476–4480, 1992
12. Yaminishi K, Inazawa J, Liew FM, Nonomura K, Ariyama T, Yasuno H, Abe T, Doi H, Hirano J, Fukushima S: Structure of the gene for human transglutaminase 1. *J Biol Chem* 267: 17858–17863, 1992
13. Kim H-C, Lewis MS, Gorman JJ, Park S-C, Girard JE, Folk JE, Chung S-I: Protransglutaminase E from guinea pig skin—isolation and partial purification. *J Biol Chem* 265:21971–21978, 1990
14. Kim I-G, Gorman JJ, Park S-C, Chung SI, Steinert PM: The deduced sequence of the novel soluble zymogen epidermal transglutamianse (TGase3) of mouse and man. *J Biol Chem* 268:12682–12690, 1993
15. Fesus L, Davies PJA, Piacentini M: Apoptosis: molecular mechanism in programmed cell death. *Eur J Cell Biol* 56:170–177, 1991
16. Knight CRL, Rees RC, Griffin M: Apoptosis: a potential role for cytosolic transglutaminase and its importance in tumour progression. *Biochim Biophys Acta* 1096:312–318, 1991
17. Haake AR, Polakowska RR: Cell death by apoptosis in epidermal biology. *J Invest Dermatol* 101:107–112, 1993
18. Hohl D: Cornified cell envelope. *Dermatologica* 180:201–211, 1990
19. Steven AC, Steven AC: Protein composition of epidermal keratinocyte cornified cell envelopes. *J Cell Sci* (in press)
20. Reichert U, Michel S, Schmidt R: The cornified cell envelope: a key structure of terminally differentiating keratinocytes. In: Darmon M, Blumenberg M (eds.). *Molecular Biology of the Skin.* Academic Press Inc., London, 1993, pp 107–150
21. Frohman MA: RACE, a rapid amplification of cDNA ends. In: Innis MA, Gelfand DH, Sninski JJ, White TJ (eds.). *PCR Protocols: A Guide to Methods and Applications.* Academic Press Inc., New York, 1990, pp 28–38

22. Devereux J, Heaberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387–395, 1984
23. Chou PY, Fasman GD: Conformation parameters for amino acids in helical, b-sheet, and random coil regions calculated from proteins. *Biochemistry* 13:222–245, 1974
24. Garnier J, Osguthorpe DJ, Robson B: Analysis of the accuracy and implications of simple methods for predicting the secondary structures of globular proteins. *J Mol Biol* 120:97–118, 1978
25. Pearson WR, Lipman DJ: Improved tools for biological structure determination. *Proc Natl Acad Sci USA* 85:2444–2448, 1988
26. Xiong Y, Eickbush TH: Organization and evolution of the retro-elements based upon their reverse trascriptase sequences. *EMBO J* 9:3353–3362, 1990
27. Korsgren C, Cohen CM: Organization of the gene for human erythrocyte membrane protein 4.2: structural similarities with the gene for the *a* subunit of factor XIII. *Proc Natl Acad Sci USA* 88:4840–4844, 1991
28. Ichinose A, Davie EW: Characterization of the gene for the a subunit of human factor XIII (plasma transglutaminase), a blood coagulation factor. *Proc Natl Acad Sci USA* 85:5829–5833, 1988
29. Mount SM: A catalog of splice junction sequences. *Nucleic Acids Res* 10:459–472, 1982
30. Markova NG, Marekov LN, Chipev CC, Gan S-Q, Idler WW, Steinert PM: Profilaggrin is a major epidermal calcium binding protein. *Mol Cell Biol* 13:613–625, 1993
31. Singer MA, Hortsch M, Goodman CS, Bentley D: Annulin, a protein expressed in limb segment boundaries in the grasshopper embryo, is homologous to protein cross-linking transglutaminases. *Dev Biol* 154:143–159, 1992
32. Chakravarty R, Rong X, Rice RH: Phorbol ester-stimulated phosphorylation of keratinocyte transglutaminase in the membrane-anchorage region. *Biochem J* 271:25–30, 1990
33. Board PG, Webb GC, McKee J, Ichinose A: The chromosomal location of the gene for human blood clotting a subunit of factor XIII. *Cytogenet Cell Genet* 48:25–27, 1988
34. Najfeld V, Ballard SG, Menninger J, Ward DC, Bouhassira EE, Swartz RS, Nagel RL, Rybicki AC: The gene for human erythrocyte protein 4.2 maps to chromosome 15q15. *Am J Human Genet* 50:71–75, 1992